Reconstruction quasi-dense et modèles 3D à partir d'une séquence d'images

Maxime Lhuillier, Long Quan

Reconstruction quasi-dense et modèles 3D à partir d'une séquence d'images

Quasi-Dense Reconstruction and 3D Models from Image Sequence

Maxime Lhuillier[1] Long Quan[2]

[1] LASMEA-UMR 6602 UBP/CNRS, 2 avenue des Landais, 63177 Aubière Cedex, France.
[2] Department of Computer Science, HKUST, Clear Water Bay, Kowloon, Hong Kong SAR.
lhuillie@lasmea.univ-bpclermont.fr quan@cs.ust.hk

Papiers et démos: wwwlasmea.univ-bpclermont.fr/Personnel/Maxime.Lhuillier

Résumé

Ce papier propose une reconstruction quasi-dense à partir d'une séquence d'images non calibrées ainsi qu'un système associé de reconstruction de modèles 3D. La principale innovation est que toute la géométrie est calculée à partir de mises en correspondances quasi-denses sous-échantillonnées au lieu des points d'intérêts épars usuels. Cela produit non seulement une reconstruction plus précise (au sens des incertitudes) et plus robuste grâce à des mises en correspondances bien redondantes et réparties dans les images, mais aussi une reconstruction plus adéquate (car plus dense) pour l'application de la reconstruction de surface. Des expériences sur des séquences réelles montrent de meilleures performances des reconstructions quasi-denses par rapport aux reconstructions éparses à la fois en robustesse et incertitudes. De plus, les surfaces de nombreux objets ont été obtenues à partir des points quasi-denses reconstruits.

Mots Clef

Vision 3D et géométrie (mise en correspondance, reconstruction 3D, évolution de surfaces, méthode des iso-surfaces, système de modélisation 3D).

Abstract

This paper proposes a quasi-dense reconstruction from uncalibrated sequence and a companion system for 3D model reconstruction. The main innovation is that all geometry is computed based on re-sampled quasi-dense correspondances rather than the standard sparse points of interest. It not only produces more accurate (according to the uncertainties) and more robust reconstruction due to highly redundant and well spread input data, but also fills the gap of insufficiency of sparse reconstruction for surface reconstruction. Experiments on real sequences demonstrates the superior performance of quasi-dense w.r.t. sparse reconstruction both in robustness and uncertainty. Also, many surfaces are calculated from quasi-dense reconstructions.

Keywords

3D Vision and Geometry (Matching, 3D Reconstruction, Surface Evolution, Level-sets, System of 3D Modeling).

1 Introduction

La reconstruction 3D à partir d'une séquence non calibrée a été un domaine très actif durant la dernière décennie en vision par ordinateur. Ceci est principalement du à la formulation intrinsèque de contraintes géométriques en géométrie projective et une meilleure compréhension des propriétés numériques et statistiques des estimations géométriques [21, 48]. Plusieurs algorithmes de reconstruction basés sur des points ont été publiés pour des séquences d'images courtes [5, 16, 18, 7] ou longues [44, 42]. La plupart de toutes ces approches sont basées sur des points d'intérêts. Des systèmes plus récents et complets sont reportés [32, 10, 37, 1, 24], pour lesquels seul une connaissance partielle des paramètres intrinsèques est nécessaire. Malheureusement, la plupart des applications de modélisation et visualisation nécessitent des reconstructions denses ou quasi-dense au lieu d'un nuage de points épars. Les méthodes denses stéréo sont limitées à des caméras précalibrées et points de vues très proches les uns des autres [43, 34, 20, 19]. On note que bien que les résultats finals reportés [36, 33] sont des modèles texturés denses, les méthodes appliquent seulement une méthode stéréo dense basée sur un algorithme de corrélation après avoir obtenu la géométrie à l'aide d'une méthode éparse.

Une approche intermédiaire entre reconstruction éparse et dense est proposée ici pour une caméra (un appareil photo numérique) tenue à la main. Par reconstruction quasi-dense, il faut comprendre que la géométrie est directement calculée sur des points sous-échantillonnés de correspondances quasi-denses, au lieu de reconstructions de points d'intérêts épars. La mise en correspondance quasi-dense est préférable à la mise en correspondance complètement dense grâce à sa plus grande robustesse et efficacité avec des caméras tenues à la main. La principale innovation est que toute la

géométrie est calculée grâce à un algorithme de mise en correspondances quasi-dense simple et efficace décrit dans [26] et appliqué au rendu basé image [25]. Les correspondances quasi-denses sont utilisées dès le début de l'algorithme d'estimation de géométrie, des modules d'estimations de géométries de 2 et 3 vues jusqu'à l'étape de fusion de sous-séquences. Ceci donne non seulement un ensemble de points reconstruits plus adéquat que quelques points épars pour calculer des surfaces, mais aussi des estimations plus précises au sens des incertitudes et plus robustes des caméras et de la structure de la scène, grâce à plus de redondance et à une répartition plus uniforme. Ce papier est composé de parties de [27] pour la reconstruction quasi-dense et de [28] pour la reconstruction de surface.

2 Mise en correspondance quasi-dense

Le calcul de la mise en correspondance quasi-dense débute par la mise en correspondance de quelques points d'intérêts appelés "germes". Un algorithme de croissance de régions propage ensuite la mise en correspondance à partir des germes obtenus, des zones les plus texturées (et donc dans lesquelles la mise en correspondance est la plus fiable) aux autres zones moins texturées.
L'algorithme [26] se décompose donc en deux étapes: la sélection des germes et la propagation, qui sont illustrés dans la Figure 1. Les points d'intérêts [29, 14] sont naturel-

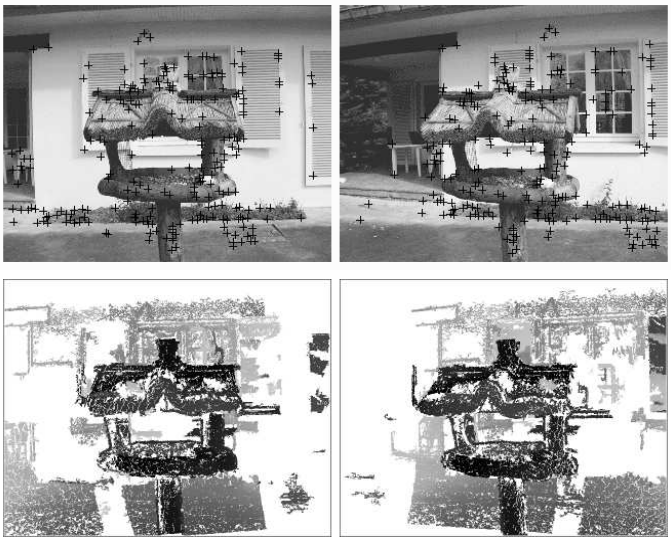

FIG. 1 – *Haut: appariements germes de deux images avec de grandes disparit´es (quelques germes sont erron´es principalement `a cause des textures p´eriodiques des volets). Bas: la propagation r´esultante sans la contrainte ´epipolaire.*

lement de bons candidats puisqu'ils sont détectés comme les points réalisant les maxima locaux de textures (au sens de Harris [14]). Les points d'intérêts sont d'abord détectés, puis mis en correspondance grâce au score de corrélation ZNCC (Zero Mean Normalized Cross Correlation, ou encore le cosinus des luminances de moyennes translatées en 0), et enfin une étape de vérification croisée est appliquée [11]. Ceci donne la liste de départ des appariements germes triée par ZNCC. Le score ZNCC est utilisé car il est plus exigeant que les scores du type valeurs absolues (ou carrés) de différences dans les régions uniformes, est plus tolérant dans les zones texturées ou le bruit est important, et est invariant aux changements de luminance affines positifs (cela permet d'encaisser plus facilement les variations de gains de caméras et parfois les surfaces non lambertiennes).
A chaque étape de la propagation, l'appariement $(\mathbf{x}, \mathbf{x}')$ composé de deux pixels correspondants $\mathbf{x}, \mathbf{x}'$ avec le meilleur ZNCC est retiré de la liste courante des appariement germes. Ensuite, les nouveaux appariements potentiels $(\mathbf{u}, \mathbf{u}')$ sont recherchés dans un voisinage spatial immédiat. Ce voisinage impose une limite de 1 sur le gradient de la disparité dans les deux dimensions des images $||(\mathbf{u}' - \mathbf{u}) - (\mathbf{x}' - \mathbf{x})||_\infty \leq 1$ pour traiter les cas où la contrainte épipolaire est inconnue ou imprécise. La contrainte d'unicité et la fin de la propagation sont garanties en considérant les pixels $\mathbf{u}, \mathbf{u}'$ qui n'ont pas encore été choisis. La taille de la fenêtre (indéformable) de corrélation est plus petite dans l'étape de propagation que dans celle de la sélection, afin d'encaisser plus facilement les distortions géométriques entre images.
La complexité en temps de cet algorithme de propagation est $O(nlog(n))$, dépend seulement du nombre d'appariements final n, et la complexité en espace est linéaire en la surface des images. Ces deux complexités sont indépendantes d'une limite sur la disparité. On note qu'à chaque instant, seul le meilleur germe est choisi, ceci limite beaucoup le risque de mauvais appariements. Par exemple, la sélection des germes est très similaire à celles déjà utilisée [49, 45] pour apparier des points par corrélation, mais la grande différence est que seuls les appariements les plus certains sont nécessaires, plutôt que d'en utiliser un maximum d'entre eux. Dans les cas extrêmes, un seul appariement germe correct suffit à provoquer une avalanche de mise en correspondance dans les zones texturées des images. Ceci rend l'algorithme moins vulnérable aux faux germes. Le même principe s'applique à la propagation, le risque de propagation erronée est réduit considérablement par la stratégie meilleur d'abord sur tout les points appariés germes.

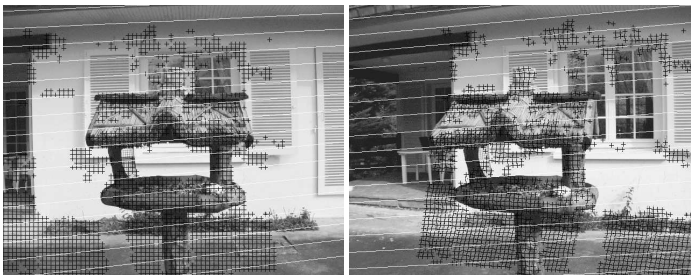

FIG. 2 – *Le sous-´echantillonnage de la mise en correspondance quasi-dense (d´ecrit en section 4) est repr´esent´es par un ensemble de croix noires dans les deux images. Ces appariements bien r´epartis sont utilis´e pour estimer la matrice fondamentale, dont certaine des lignes ´epipolaires sont dessin´ees.*

3 Sous-échantillonnage

L'ensemble des mises en correspondances obtenu tel quel n'est pas approprié pour estimer la géométrie car le coût d'un calcul direct serait très lourd à cause du grand nombre de mises en correspondances. La méthode de sous-échantillonnage proposée ici produit non seulement un ensemble réduit, plus uniformément distribués de points appariés dans les images que des points d'intérêts appariés (une répartition uniforme est recommendée pour les calculs géométriques [17]), mais aussi des points à une précision sous-pixelique. Cet étape de sous-échantillonnage est également motivée par un besoin de régularisation pour améliorer la fiabilité de la mise en correspondance en intégrant une contrainte géométrique locale, ceci afin d'éliminer des appariements erronés.

En supposant que la surface de la scène observée est suffisamment lisse pour pouvoir être approchée par des morceaux de surface plans, on estime alors localement des homographies à partir de la mise en correspondance. La première image est subdivisée à l'aide d'une grille régulière en petites portions carrées. Pour chaque portion, on considère tous les pixels appariés dans le carré provenant de l'appariement quasi-dense. On estime alors de façon robuste avec la méthode RANSAC [8] une homographie à partir des appariements dans ce carré afin d'obtenir une portion de surface potentielle. Le compromis choisi entre la dimension de la portion et la stabilité de l'estimation est d'estimer une application affine (qui comptabilise seulement 6 degrés de libertés contre 8 pour une homographie générale) dans des portions 8×8 pixels. Finalement, pour chaque portion confirmée par une homographie estimée $\mathbf{H}_i$, un seul appariement à précision sous-pixelique $(\mathbf{u}_i, \mathbf{H}_i\mathbf{u}_i)$ est retenu pour la suite de la méthode de reconstruction proposée. Ces appariements sont illustrés par des croix en figure 2, qui sont mieux répartis dans l'espace image que des points d'intérêts montrés en haut de la figure 1.

4 Estimer la géométrie de 2 vues

La géométrie de 2 vues d'une scène rigide est décrite par la matrice fondamentale. L'approche standard habituelle consiste à calculer automatiquement et simultanément la matrice fondamentale et les mises en correspondances éparses de points d'intérêts [49, 45]. Il y a aussi des tentatives pour intégrer le calcul d'une mise en correspondance dense dans l'optimisation non linéaire de la matrice fondamentale en minimisant un score de corrélation global [12], mais l'algorithme est très coûteux en calcul (7-12 minutes contre 20-40 secondes pour la méthode proposée ici pour des images 512×512 sur des machines similaires) et ne traite pas les zones partiellement occultées comme celles des images distantes montrées en Figure 2. Dans le contexte de l'algorithme de mise en correspondance quasi-dense, il y a deux façons d'intégrer l'estimation de géométrie. La première est une propagation contrainte par la matrice fondamentale (estimée par la mise en correspondance éparse initiale), et la seconde une propagation non contrainte. L'avantage de la propagation contrainte est que les propagations erronnées sont fort probablement stoppées plus tôt, mais le domaine de propagation dans les images peut être réduit. Plus grave, la géométrie estimée avec une méthode robuste s'estime parfois dans un sous-ensemble seulement des images (par exemple, l'arrière ou l'avant plan), ce qui ne conduit pas à la distributions uniforme recommendée pour les calculs d'estimations géométriques [17]. On préfère donc une stratégie de propagation non contrainte suivie par une propagation contrainte, plus robuste que la première, comme indiqué ci-dessous.

1. Détecter les points d'intérêts dans deux images et calculer les premiers appariements par corrélation et vérification croisée
2. Effectuer une propagation non contrainte à partir de ces points
3. Sous-échantillonner la mise en correspondance quasi-dense et lister les appariements résultants dans $\mathbf{L}_1$
4. Apparier les points d'intérêts grâce aux homographies estimées dans l'étape précédante, et rajouter les appariements résultants à $\mathbf{L}_1$
5. Estimer la matrice fondamentale $\mathbf{F}$ à l'aide d'un algorithme robuste standard [45, 49, 17] à partir de $\mathbf{L}_1$
6. Effectuer une propagation contrainte par $\mathbf{F}$
7. Sous-échantillonner la mise en correspondance quasi-dense dans $\mathbf{L}_2$, et compléter $\mathbf{L}_2$ avec les points d'intérêts appariés grâce aux homographies estimées
8. Estimer la matrice fondamentale $\mathbf{F}$ à l'aide d'un algorithme robuste standard [45, 49, 17] à partir de $\mathbf{L}_2$

5 Estimer la géométrie de 3 vues

La géométrie de 3 vues joue un rôle central pour la construction de séquences longues : au plus 3 caméras se calculent explicitement et il faut des contraintes sur 3 caméras pour éliminer l'ambiguité de la mise en correspondance. La reconstruction projective pour 3 caméras avec un minimum de 6 points est l'outil de calcul clef utilisé pour valider les appariements à l'aide de la méthode robuste RANSAC et pour initialiser un ajustement de faisceaux [38, 39, 46, 17, 40]. L'algorithme quasi-dense 3 vues est le suivant :

1. Appliquer l'algorithme quasi-dense 2 vues entres les images i et $i-1$, et entre i et $i+1$
2. Calculer la liste d'appariements tous sur 3 vues $i-1, i, i+1$ en fusionnant les deux listes d'appariements sous-échantillonnées provenant de la paire $i, i-1$ et de la paire $i, i+1$
3. Appliquer RANSAC en tirant aléatoirement 6 appariements sur 3 vues pour calculer explicitement 3 caméras dans une base projective [38] (algorithme dit "des 6 points") et rejeter les appariements erronés en fonction de leurs erreurs de reprojection. Les triplets sont reconstruits connaissant les caméras et puis reprojetés pour calculer leurs erreurs de reprojection.

4. Appliquer un ajustement de faisceaux sur la géométrie des 3 vues $i-1, i, i+1$ avec tous les triplets non rejetés, en minimisant la somme de leurs erreurs de reprojections.

Le principe qui consiste à estimer la géométrie des triplets de vues consécutives pour reconstruire une séquence d'images longue a déjà été appliqué par [10, 24, 46], et voici des différences avec l'approche présentée ici :

- Les points appariés par paires ne sont pas transférés pour de la mise en correspondance guidée sur 3 vues. Ici, la géométrie 3 vues ne sert qu'à valider ou rejeter les mises en correspondances sous-échantillonnées sur 3 vues, et le pourcentage de rejet est faible.
- La paramétrisation du tenseur trifocal pour la géométrie n'est pas utilisée comme cela est suggéré par [10, 1, 17, 40]. Les 3 matrice caméras dans une base projective se calculent directement par l'algorithme des 6 points, et sont utilisées pour reconstruire les points et calculer les erreurs de reprojection.

  La paramétrisation par tenseurs est peu justifiable ici car cela donnerait une sur-paramétrisation compliquée de la géométrie de 3 vues, et des algorithmes numériques plus sophistiqués seraient nécessaires pour son estimation. Les tenseurs pourraient être utiles pour faire de la mise en correspondance guidée [9], mais ceci n'est pas nécessaire ici.

6 Fusion hiérarchique

La stratégie de fusion hiérarchique projective utilisée par [10, 24] est adaptée pour calculer une séquence entière à partir des paires et triplets quasi-denses, ce qui est plus efficace qu'une stratégie de fusion incrémentale simple. L'algorithme quasi-dense de fusion hiérarchique est le suivant :

1. Pour chaque paire d'images consécutives dans la séquence, appliquer l'algorithme quasi-dense 2 vues.
2. Pour chaque triplets d'images consécutives dans la séquence, appliquer l'algorithme quasi-dense 3 vues.
3. L'approche est hiérarchique : une séquence longue $[i..j]$ est obtenue en fusionnant $[i..k+1]$ et $[k..j]$ avec deux vues k et $k+1$ communes, et k est l'image d'index milieu de la séquence $[i..j]$. La fusion consiste à
   (a) Fusionner les mises en correspondance sous-échantillonnées des deux sous-séquences sachant que les images k et $k+1$ sont communes
   (b) Estimer par moindres carrés linéaires l'homographie de l'espace qui effectue le changement de base projectif entre les deux sous-séquences pour les caméras k et $k+1$
   (c) Appliquer cette homographie sur les caméras et points propres à une sous-séquence afin de tout exprimer dans une même base projective
   (d) Effectuer un ajustement de faisceaux sur la séquence $[i..j]$ avec tous les points fusionnés

Plusieurs algorithmes ont étés proposés [10] pour fusionner deux séquences de 3 vues avec 0, 1, ou 2 vues communes. Le principal avantage d'avoir 2 vues communes est que les matrices caméras suffisent pour estimer l'homographie reliant les deux sous-séquences : choisir des points communs est inutile. Il est aussi important de remarquer qu'à la fois des points d'intérêts et des point "généraux" mis en correspondances (i.e. qui proviennent de sous-échantillonnage de mise en correspondance quasi-dense) interviennent dans tout le calcul.

7 Estimation euclidienne optimale

L'étape finale de la reconstruction quasi-dense consiste à passer d'une reconstruction projective à une reconstruction métrique en appliquant une méthode d'auto-calibration puis une optimisation de la géométrie entière.

- Une méthode linéaire [36] basée sur la paramétrisation de la quadrique absolue [47] est utilisée pour estimer la distance focale, inconnue et identique pour toutes les caméras, les autres paramètres intrinsèques étant connus.
- Cette méthode donne aussi l'homographie de l'espace qui permet de passer d'une reconstruction projective à une reconstruction métrique. Le système de coordonnées métrique choisi est celui de la caméra d'index milieu et d'échelle telle que la distance maximale entre deux positions de caméras est 1.
- Un ajustement de faisceaux est ensuite appliqué sur toutes les caméras et tous les points quasi-denses. Les caméras sont paramétrées par leurs 6 paramètres extrinsèques et une distance focale commune à toutes les caméras. Cette paramétrisation naturelle permet de traiter toutes les caméras de la même façon.
- Un second ajustement de faisceaux euclidien incluant un paramètre de distortion radiale commun à toutes les caméras est appliqué lorsque celle-ci est non négligeable, notamment pour les séquences d'images capturées avec une distance focale courte.

Il est clair que l'implémentation des ajustements de faisceaux projectifs et euclidiens exploite la structure creuse des systèmes linéaires à résoudre, comme cela est suggéré en photogrammétrie [2, 41] et en vision [15, 48, 17], puisque le nombre de points 3D quasi-dense dépasse fréquemment 20000 (pour cela, les systèmes creux sont réduits à des sous-systèmes denses en les paramètres des caméras, en éliminant les paramètres de structure).
On note enfin que tous les ajustements de faisceaux sont appliqués plusieurs fois, ceci afin d'augmenter entre deux le nombre des points consistants avec la géométrie, et donc d'améliorer leur distribution dans les images.

8 Comparaison EPARS / QUASI

La précision au sens des incertitudes et la robustesse sont comparées entre la méthode de reconstruction quasi-dense

(QUASI) et la méthode classique éparse (EPARS). En fait, deux algorithmes de reconstructions épars basés sur des points d'intérêts sont considérés. Le premier consiste simplement à tracker tous les points d'intérêts détéctés dans les images. Le second est une mixture éparse et quasi-dense : il consiste à sélectionner les mises en correspondances de points d'intérets compatibles avec la géométrie calculée avec la mise en correspondance quasi-dense, et à re-évaluer la géométrie complète seulement à partir de ces points d'intérêts sélectionnés. Dans la suite, on désigne par EPARS le meilleur résultats de ces deux méthodes.

Pour mesurer la précision de la reconstruction, on pourrait considérer le résultat de l'ajustement de faisceau comme un estimateur au maximum de vraissemblance des paramètres des caméras et de la structure de la scène, si on admet que les points images sont normalement distribués autour de leur positions exactes avec un écart type inconnu σ, ce qui semble raisonnable [22].

La matrice de covariance est seulement définie au choix de gauge près [48, 30, 31] et l'échelle de bruit près σ^2. Le niveau de bruit σ^2 est estimé à partir des erreurs de reprojection par $\sigma^2 = r^2/(2e - d)$ avec r^2 est la somme des e erreurs de reprojection élévées au carré, d est nombre de paramètres indépendants de la minimisation $d = 1 + 6c + 3p - 7$ (1 pour la distance focale commune, c le nombre de caméras, p le nombre de points et 7 le nombre de degrés de libertés pour le choix de gauge, c'est à dire la dimension du groupe des similitudes). Tous les résultats donnés ici sont "gauge free" : la matrice de covariance est calculée sans imposer de contrainte pour le choix de la gauge, maintenant dans le système de coordonnées de la caméra d'index milieu de la séquence et avec l'échelle telle que la distance maximale entre deux centre de caméras soit 1. Les conclusions sont les mêmes pour la comparaison entre EPARS et QUASI avec un choix (partiel) de gauge "basée caméra" en fixant l'orientation et la position de la caméra d'index milieu (en particulier, l'incertitude σ_f de la distance focale commune f est la même pour tout les choix de gauge car f est gauge-invariant). Puisque la matrice de covariance est de dimension très grande, seuls ses blocs diagonaux pour les caméras et les points sont calculés en utilisant une méthode de pseudo-inversion creuse [48, 17, 31].

Des ellipsoides de confiance à 90% pour chaque position 3D sont choisis : si $\mathbf{C}$ est la matrice de covariance d'une position de caméra ou d'un point extraite de la matrice de covariance exacte de tout les paramètres, l'ellipsoide de confiance est alors défini par $\Delta\mathbf{x}^T\mathbf{C}^{-1}\Delta\mathbf{x} \leq 6.25$ (ellipsoide de confiance pour une probabilité de 90% et une loi du Khi-deux à 3 degrés de libertés). Le plus grand des demi-axes des ellipsoides à 90% est considéré ici comme une borne pour chaque position 3D,et donc comme une précision. Comme le nombre de caméras est faible, on évalue simplement la précision de la position des centres des caméras comme la moyenne de leur borne $\overline{\mathbf{x}}_{c_i}$. Le nombre de point est lui vraiment important, particulièrement pour la méthode QUASI. Pour avoir une évaluation plus détaillée de la précision des points 3D, on calcule les bornes quantiles à 0% (la plus petite borne $\mathbf{x}_0$), à 25% ($\mathbf{x}_{\frac{1}{4}}$), à 50% (la médiane des bornes $\mathbf{x}_{\frac{1}{2}}$), à 75% ($\mathbf{x}_{\frac{3}{4}}$) et à 100% (la plus grande borne $\mathbf{x}_1$). Notons enfin que l'on a supposé que l'estimation obtenue de la géométrie est suffisament proche de la géométrie exacte afin de pouvoir approximer la matrice de covariance exacte par la matrice de covariance effectivement calculée, et que la justification du calcul de la matrice de covariance exacte nécessite elle-même de nombreuses approximations.

8.1 Exemples

Des résultats expérimentaux sont donnés sur trois séquences réelles. La séquence Corridor (11 images 512×512) a un mouvement vers l'avant le long de la scène et favorise la méthode EPARS car c'est une scène polyèdrique peu texturée, les points d'intérêts appariés sont abondants et bien répartis dans les images. La séquence Lady (20 images 768×512) a un mouvement latéral plus favorable pour estimer la profondeur de la scène.

Séquence Corridor La table 2 compare les évaluations de la précision au sens des incertitudes pour la séquence Corridor. Avec 40 fois plus de redondance, la position des caméras (resp. la distance focale) avec QUASI sont 2 fois (resp. 4 fois) meilleurs que ceux de EPARS. Cependant, l'évaluation de la précision des points pour EPARS est meilleure que ceux de QUASI pour la majorité des points. Comme la direction du mouvement de la caméra est voisine de celle de Corridor, les points au fond du Corridor ont une précision relativement mauvaise par rapport aux autres, ce à quoi on pouvait s'attendre. La figure 3 montre les résultats de reconstruction obtenus avec pour chaque point 3D une petite portion texturée dans son voisinage, et une projection sur un plan horizontal des ellipsoides de confiance à 90%.

Séquence Lady Pour la séquence Lady, les résultats obtenus par les méthodes QUASI et EPARS sont dans la table 2 et la figure 4. L'évaluation de la précision pour QUASI est meilleure que pour EPARS, 6 fois meilleure pour la distance focale et la position des caméras. Les conclusions pour les séquences de la figure 5 sont similaires à celles de Lady lorsque EPARS réussit. On a noté que le faible nombre de points 3D EPARS sur cet exemple rend l'estimation EPARS peu robuste.
Enfin, les temps de calculs en minutes pour la méthode QUASI sont donnés à droite de la table 2 sur un P4 à 2.4Ghz.

8.2 Robustesse

On mesure simplement la robustesse relative des méthodes EPARS et QUASI en regardant leurs cas d'échecs (cf. la table 1) sur les séquences utilisées dans la figure 5. La méthode QUASI est plus robuste pour toutes les séquences testées, y compris toutes celles non référencées dans ce papier : si une séquence est reconstruite avec succès par EPARS alors c'est aussi le cas avec QUASI ; de plus la méthode EPARS échoue pour plusieurs autres séquences pour lesquelles la méthode QUASI réussit.

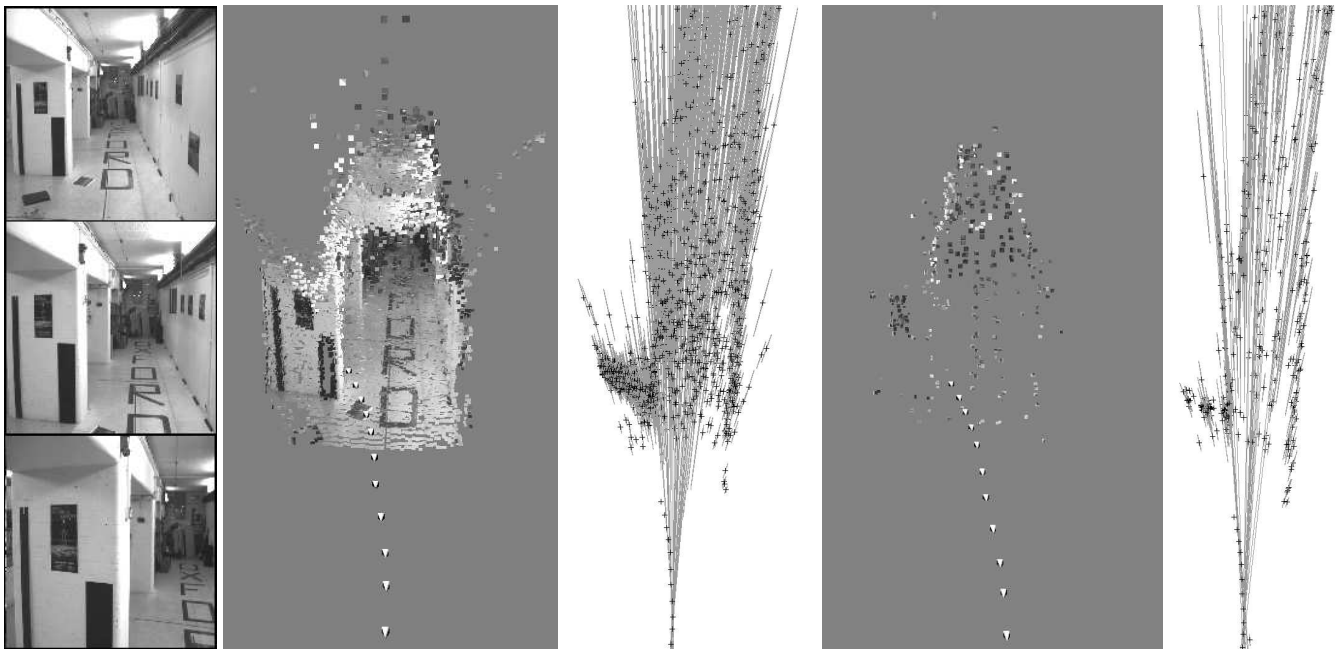

FIG. 3 – *Reconstructions QUASI (gauche) et EPARS (droite) pour Corridor et leurs ellipsoides de confi ance `a 90% projet´ees sur un plan horizontal. Seul 1 ellipsoide sur 10 est affi ch´e pour QUASI. Trois images de la s´equence Corridor sont aussi donn´ees.*

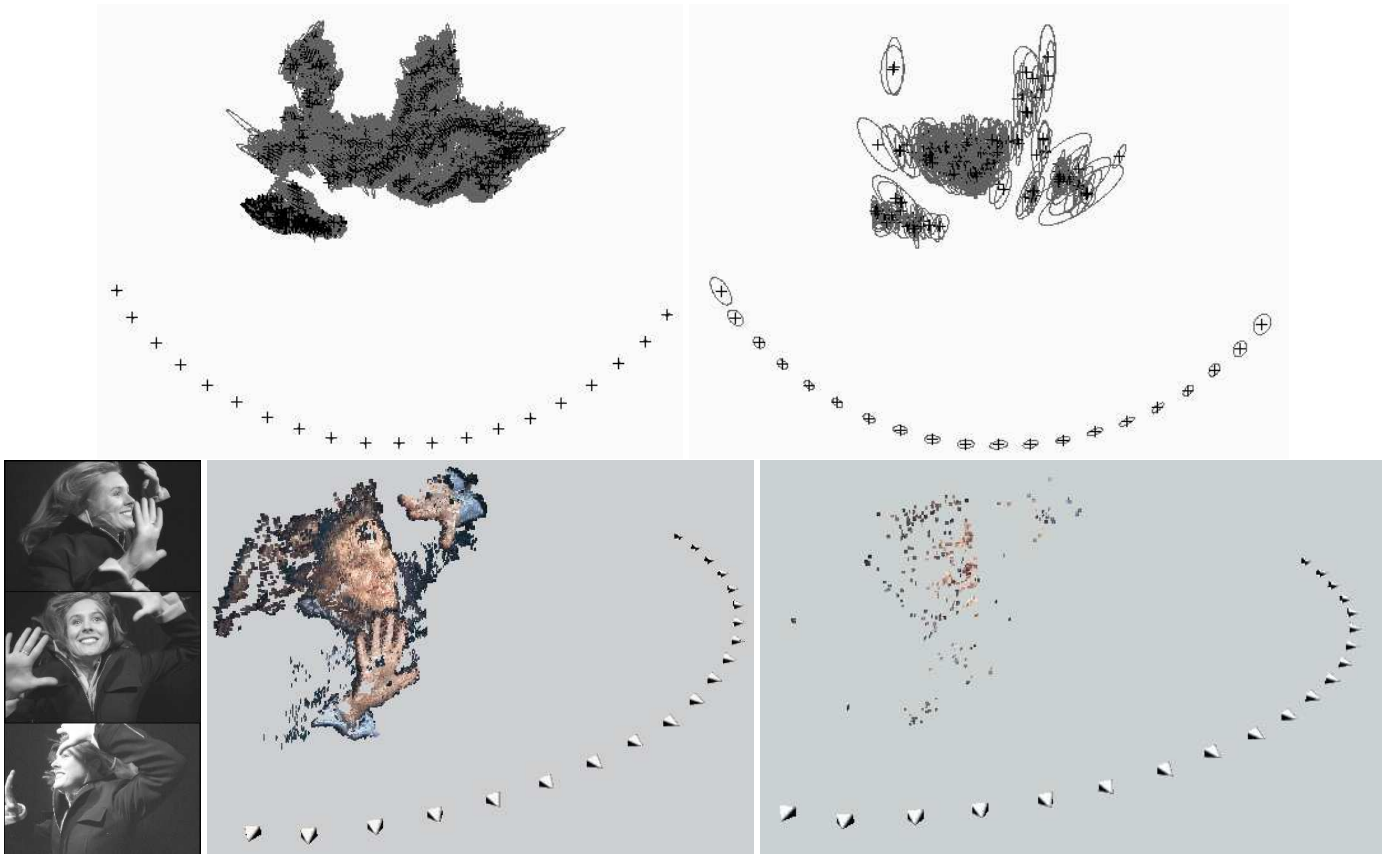

FIG. 4 – *Reconstructions QUASI (gauche) et EPARS (droite) pour Lady. Les ellipsoides `a 90% pour l'ajustement de faisceaux fi nal sont zoom´es 4 fois pour aider la visualisation. Trois images de la s´equence Lady sont aussi donn´ees.*

	Corridor	Bust	Man 1	Man 2	Lady	Girl
Q	ok	ok	ok	ok	ok	ok
E	ok	ok	rate	rate	ok	rate

TAB. 1 – *Succ`es et ´echecs des m´ethodes Q(uasi) et E(pars).*

9 Reconstruction de surface

En plus des avantages de précision au sens des incertitudes et de robustesse décrits précédemment, un grand intérêt de la reconstruction quasi-dense est que sa qualité autorise l'estimation de surfaces d'objets 3D quitte à faire un peu attention à la prise de vues (voir la section 10), ce qui est très intéressant pour les applications de synthèse d'images et de visualisation à partir de matériel grand public. Cette section résume une partie de la publication [28], les résultats et explications données ici ne concernant que la méthode qui utilise seulement les points 3D fournis par la reconstruction quasi-dense.

9.1 Evolution de surface

Brièvement, la méthodologie suivie est une approche variationnelle inspirée par plusieurs travaux [23, 4, 3, 13, 6, 50]. Dans tous ces cas, la surface S recherchée doit minimiser l'intégrale sur sa surface d'une quantité $p(S) = \int\int w ds$, avec w fonction d'un gradient d'image [23, 3], ou bien la distance à un ensemble de points 3D obtenus par scanner [50], ou encore un score de corrélation [13, 6]. Les solutions de la fonctionnelle à minimiser sont données par les équations d'Euler-Lagrange notées $\nabla p = 0$ et vérifiant $p(S+\epsilon) = p(S) + \int\int \epsilon \nabla p ds + o(\epsilon)$.
Une première façon pour résoudre $\nabla p = 0$ consisterait à déformer une surface paramétrée

$$\mathbf{x}(t) : (u, v, t) \mapsto (x(u, v, t), y(u, v, t), z(u, v, t))$$

dans le temps avec la vitesse $-\nabla p$, d'où $\mathbf{x}_t = \frac{\partial \mathbf{x}(u,v,t)}{\partial t} = -\nabla p$. Ceci est la formulation Lagrangienne du problème qui indique comment chaque point de la surface doit se déplacer pour minimiser la fonctionnelle.
Pour éviter d'avoir à gérer les changements de paramétrage 2D de la surface dans cette formulation (délicat par exemple en cas de changement de topologie), et en vérifiant que la vitesse $-\nabla p$ ne dépend pas du choix de paramétrage, on peut définir la surface déformable implicitement par $u(t, \mathbf{x}) = 0$ qui évolue dans R^3 selon $u_t = -(\nabla p \cdot \mathbf{n})||\nabla u||_2$ avec la normale définie par $\mathbf{n} = -\frac{\nabla u}{||\nabla u||_2}$, soit encore

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + w||\nabla u||_2 H$$

avec $H = div \frac{\nabla u}{||\nabla u||_2}$ la somme des deux courbures principales. Les changements de topologie, la précision et la stabilité de l'évolution sont gérées proprement par des schémas de différences finies [35] du type $u^{n+1}_{ijk} = \Delta t.f(\Delta x, u^n_{i'j'k'})$ avec u^n_{ijk} la discrétisation de u au point ijk et au temps n.

9.2 Régularisation bornée et accélération

Ici $w = w(x) = d(x, \mathcal{P})$ est la distance Euclidienne entre un point x et l'ensemble $\mathcal{P}$ des points précédemment reconstruits. En pratique, le terme de régularisation $w||\nabla u||_2 H$ lisse trop les détails de la surface et cause une convergence très lente car les pas de temps et espace doivent vérifier $\Delta t = O(\Delta x^2)$ pour une évolution stable. D'un autre coté, il est possible d'éliminer purement et simplement ce terme dans le cas de points 3D de meilleur qualité obtenus par un scanner [51], ce qui accèlere grandement l'évolution puisque la condition de stabilité devient $\Delta t = O(\Delta x)$. Une régularisation "bornée" $min(w, w_{max})||\nabla u||_2 H$ est ici proposée [28], au lieu du terme de régularisation "complet" $w||\nabla u||_2 H$. L'équation d'évolution devient alors

$$\frac{\partial u}{\partial t} = \nabla w \nabla u + min(w, w_{max})||\nabla u||_2 H.$$

On fait alors les remarques suivantes:

- l'évolution de surface avec régularisation "complète" est obtenue avec $w_{max} \geq ||w||_\infty$
- l'évolution de surface sans régularisation est obtenue avec $w_{max} = 0$.
- Lorsque $0 \approx w \leq w_{max}$ au voisinage de la surface à l'équilibre, l'évolution de surface est similaire à celle avec régularisation "complète"

Une preuve est aussi proposée dans [28] pour la condition de stabilité. On suppose pour cela que l'on impose classiquement $||\nabla u||_2 = 1$ pour éviter les trop grandes et trop faibles variations de u. On a alors $||\nabla u||_2 div \frac{\nabla u}{||\nabla u||_2} = \triangle u$, et sous cette hypothèse, on montre que la stabilité $||u^{n+1}||_\infty \leq ||u^n||_\infty$ est vérifiée si $\Delta t \leq \Delta t_{max}$ avec

$$\Delta t_{max} = \frac{\Delta x^2}{6 w_{max} + ||\Delta x(|d^{0x} w| + |d^{0y} w| + |d^{0z} w|)||_\infty}.$$

pour la discrétisation centrée de $\triangle u$ et "up-wind" de $\nabla w \nabla u$.

9.3 Calcul de la surface initiale

Avant le calcul de surface initiale, les points 3D reconstruits sont segmentés entre l'objet à l'avant plan et l'arrière plan. L'arrière plan inclus les points erronés évidents de l'avant plan que sont les points isolés distants de la majorité des autres points. Plus précisément, les points à l'avant plan sont obtenus comme la plus grande composante connexe du graphe de voisinage de tous les points telle que deux points forment une arête du graphe si leur distance est inférieur à un multiple de l'incertitude médiane des points.
La surface initiale est alors obtenue de la façon suivante. Les points segmentés à l'avant plan sont régulièrement séparés en tranches perpendiculaires à l'axe principal du nuage de point. Une enveloppe convexe 2D est calculée pour chaque tranche, et ces enveloppes convexes sont utilisées pour définir les sections successives d'un volume englobant de l'objet.
On précise aussi que tous les points 3D retenus pour le calcul de la surface sont étalés au mieux dans un ensemble de $150 \times 150 \times 150$ voxels en y appliquant une similitude. La taille des voxels alors obtenus est du même ordre de grandeur que la médiane des incertitudes des points.

10 Conditions expérimentales et limites

Voici des conditions expérimentales à respecter et des limitations pour chacune des étapes suivantes.

10.1 Estimation robuste projective

L'algorithme de reconstruction quasi-dense proposé ici s'applique à des séquences d'images non calibrées ordonnées, pour laquelle une unique matrice fondamentale doit exister entre deux images consécutives (sinon la solution au problème n'est pas unique). La scène observée ne doit donc pas être plane, et le mouvement entre deux vues consécutives doit être suffisament grand. Pour faire le tour complet d'un objet, environ 30-35 vues sont utilisées.

Biensur, les images doivent contenir suffisamment d'information de texture, pour obtenir un ensemble de points appariés quasi-dense assez uniforme dans les images.

On a noté qu'il est souvent possible de traiter le cas des textures réputées difficiles des cheveux avec une lumière ambiante suffisamment forte et sans images floues, mais aussi les changements de gains de caméras et parfois les surfaces non lambertiennes grâce au score de corrélation utilisé invariant aux changements affines de luminances.

Enfin, l'utilisation de fenêtre de corrélation de petite taille dans l'étape de propagation permet d'encaisser plus facilement les distortions géométriques entre images.

10.2 Passage projectif-euclidien

Il existe des mouvements dit "critiques", pour lesquels plusieurs distances focales sont possibles pour un même mouvement de caméra projective (i.e. la solution au problème n'est pas unique). On remédie assez facilement en demandant à l'utilisateur une distance focale à imposer, mais ceci est rarement nécessaire en pratique, en tout cas pour l'application de la synthèse d'images.

10.3 Reconstruction de surfaces

Une méthode d'évolution de surface est utilisée, notamment pour boucher naturellement les "trous" de la surface, c'est à dire les zones sans points 3D reconstruits. Pour la méthode décrite ici, il faut cependant éviter les trous trop grands qui correspondent à des zones convexes des objets. Il faut notamment éviter les joues entièrement à l'ombre des personnages, par exemple grâce à une bonne lumière ambiante (les taches de rousseurs ne sont pas nécessaires).

Si le trou est vraiment très grand, la surface peut converger vers la surface vide. On renonce alors à reconstruire cette zone et l'utilisateur introduit manuellement un plan séparant deux demi-espaces : d'un coté la surface évolue normalement et est conservée pour le modèle 3D final, de l'autre l'évolution est gelée et la surface ignorée (souvent en bas des objets). La méthode de reconstruction de surface n'est donc pas complètement automatique.

Il faut biensur veiller à choisir les vues de manière à reconstruire toute la surface. Chaque zone de l'objet à reconstruire doit être visible dans au moins 3 images consécutives.

10.4 Fusion de texture multi-vues

L'algorithme de fusion de texture utilisé et non décrit ici est un placage de texture sur un cone, voisin du placage de texture sur un cylindre. Pour cette étape (et seulement cette étape), on fait l'hypothèse forte que le mouvement de la caméra est à peu près horizontal, que l'angle de vue fait un angle grossièrement constant avec la verticale, et pointe approximativement vers le centre de l'objet.

11 Conclusion

Dans ce papier, une méthode générale de reconstruction 3D quasi-dense à partir d'une séquence non calibrée est proposée. L'idée principale est que toute la géométrie est calculée à partir d'un sous-échantillonnage d'une mise en correspondance quasi-dense au lieu des seuls points d'intérêts appariés. Les performances obtenues en précision au sens des incertitudes et en robustesse sont meilleures grâce à une distribution des mises en correspondance plus uniforme et plus redondante dans les images. Par sa densité de points, la reconstruction quasi-dense est aussi plus applicable à la visualisation qu'une reconstruction de points d'intérêts.

On note aussi que la majorité des modèles 3D obtenus sont des visages car ceux-ci vérifient bien les hypothèses de surfaces fermées et lisses de l'étape de reconstruction de surface, bien que l'étape d'estimation géométrique ait un domaine d'application bien plus large. Le système complet constitue en lui même une avancée pratique importante car aucun autre système n'a pu être capable jusqu'à présent de sortir des modèles 3D pour des visages complets (cheveux inclus) avec du matériel et un protocole expérimental aussi simple, et sans supposer que l'objet est un visage ou que le fond est fixé ou segmenté dans les images.

Il y a plusieurs pistes pour des travaux futurs, dont la réduction du temps de calcul pour les séquences longues en choisissant astucieusement les points reconstruits dans la fusion hiérarchique, et la reconstruction de scènes extérieures vues par un piéton ou une automobile.

Références

[1] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. *ECCV'96*.

[2] D.C. Brown. The bundle adjustment – progress and prospects. *International Archive of Photogrammetry*, 21, 1976. Update of "Evolution, Application and Potential of the Bundle Method of Photogrammetric Triangulation", ISP Symposium Commission III, Stuttgart, 1974.

[3] V. Caselles, R. Kimmel . Minimal surfaces based object segmentation. IEEE *TPAMI*, 19(4):394–398, 1997.

[4] V. Caselles, R. Kimmel and G. Sapiro. Geodesic active contours. *IJCV*, 22(1):61–79, 1997.

[5] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? *ECCV'92*.

[6] O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. *ECCV'98*.

[7] O. Faugeras and Q.T. Luong. *The Geometry of Multiple Images*. The MIT Press, 2001.

[8] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fi tting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, June 1981.

[9] A.W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3d model construction for turn-table sequences. *SMILE'98*.

[10] A.W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. *ECCV'98*.

[11] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve discontinuities. *IJCAI'91*.

[12] C. Gauclin and T. Papadopoulos. Fundamental matrix estimation driven by stereo-correlation. *ACCV'00*.

[13] J. Gomes and O. Faugeras. Pde-based stereo applied to human faces captures. Esprit, improofs 23-515 project, Inria, Sophia-Antipolis, 1998.

[14] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[15] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, October 1993.

[16] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *CVPR'92*.

[17] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.

[18] A. Heyden. *Geometry and Algebra of Multiple Projective Transformations*. PhD thesis, Lund Institute of Technology, 1995.

[19] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. *ICCV'95*.

[20] T. Kanade, P.J. Narayanan, and P.W. Rander. Virtualized reality: Concepts and early results. In *Workshop on Representation of Visual Scenes, Cambridge, Massachusetts, USA*, pages 69–76, June 1995.

[21] K. Kanatani. *Statistical Optimisation for Geometric Computation: Theory and Practice*. Elsevier Science, 1996.

[22] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? *ICCV'01*.

[23] S. Kichenassamy, P. Olver, A. Kumar and A. Tannenbaum. Gradient flows and geometric active contour models. *ICCV'95*.

[24] S. Laveau. *G´eom´etrie d'un syst`eme de N cam´eras. Th´eorie, estimation, et applications*. Th`ese de doctorat,École Polytechnique, May 1996.

[25] M. Lhuillier. Mod´elisation pour la synth`ese d'images `a partir d'images. Th`ese de doctorat, INPG, December 2000.

[26] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. IEEE *TPAMI*, 24(8):1140–1146, 2002.

[27] M. Lhuillier and L. Quan. Quasi-dense reconstruction from image sequence. *ECCV'02*.

[28] M. Lhuillier and L. Quan. Surface reconstruction by integrating 3d and 2d data of multiple views. *ICCV'03*.

[29] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI'81*.

[30] P. F. McLauchlan. Gauge independence in optimization algorithms for 3D vision. *LNCS vol. 1883*, 2000.

[31] D.D. Morris, K. Kanatani and T. Kanade. Uncertainty modeling for optimal structure from motion. *LNCS vol. 1883*.

[32] D. Nister. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. *ECCV'00*.

[33] D. Nister. *Automatic Dense Reconstruction from Uncalibrated Video Sequences*. PhD thesis, Ericsson and University of Stockholms, 2001.

[34] Y. Ohta and T. Kanade. Stereo by intra and inter-scanline search using dynamic programming. IEEE *TPAMI*, 7(2):139–154, 1985.

[35] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[36] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *ICCV'98*.

[37] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Metric 3d surface reconstruction from uncalbrated image sequences. *SMILE'98*.

[38] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. IEEE *TPAMI*, 17(1):34–46, January 1995.

[39] F. Schaffalitzky, A. Zisserman, R. Hartley, and P.H.S. Torr. A six point solution for structure and motion. *ECCV'00*.

[40] A. Shashua. Trilinearity in visual recognition by alignment. *ECCV'94*.

[41] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[42] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. *ECCV'96*.

[43] H. Tao, H.S. Sawhney and R. Kumar. A global matching framework for stereo computation. *ICCV'01*.

[44] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9(2):137–154, November 1992.

[45] P.H.S. Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, 1997.

[46] P.H.S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *IVC*, 5(15):591–605, 1997.

[47] B. Triggs. Autocalibration and the absolute quadric. *CVPR'97*.

[48] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle ajustment - a modern synthesis. *LNCS vol. 1883*.

[49] Z. Zhang, R. Deriche, O.D. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI*, 78(1-2):87–119, 1995.

[50] H.K. Zhao, B. Merrigan, S. Osher and M. Kang. Implicit and non-parametric shape reconstruction from unorganized points using variational level set method. *CVIU*, 80:295–319, 2000.

[51] H.K. Zhao, S. Osher and R. Fedkiw. Fast surface reconstruction using the level set method. In *IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 2001.

Corridor	#3D points	σ	f	σ_f	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	2 vues	3-n vues
QUASI	16976	0.41	714	4.36	7.0e-4	.014	.070	.13	.38	15700	2 min.	4 min.
EPARS	427	0.52	761	17.3	1.7e-3	.016	.056	.12	.32	106	-	-
Lady	#3D points	σ	f	σ_f	$\overline{\mathbf{x}}_{c_i}$	$\mathbf{x}_0$	$\mathbf{x}_{\frac{1}{4}}$	$\mathbf{x}_{\frac{1}{2}}$	$\mathbf{x}_{\frac{3}{4}}$	$\mathbf{x}_1$	2 vues	3-n vues
QUASI	26823	0.53	849	2.26	6.1e-4	1.2e-3	4.2e-3	5.4e-3	6.1e-3	1.9e-2	4 min.	6 min.
EPARS	383	0.54	866	13.6	3.8e-3	5.2e-3	9.8e-3	1.1e-2	1.3e-2	2.6e-2	-	-

TAB. 2 – *Gauche: ´evaluation de la pr´ecision au sens des incertitudes pour les s´equences Corridor et Lady (la moyenne des bornes d'incertitudes pour les centre des cam´eras et plusieurs quantiles des bornes pour les points 3D). Droite: temps de calculs sur P4 2.4Ghz.*

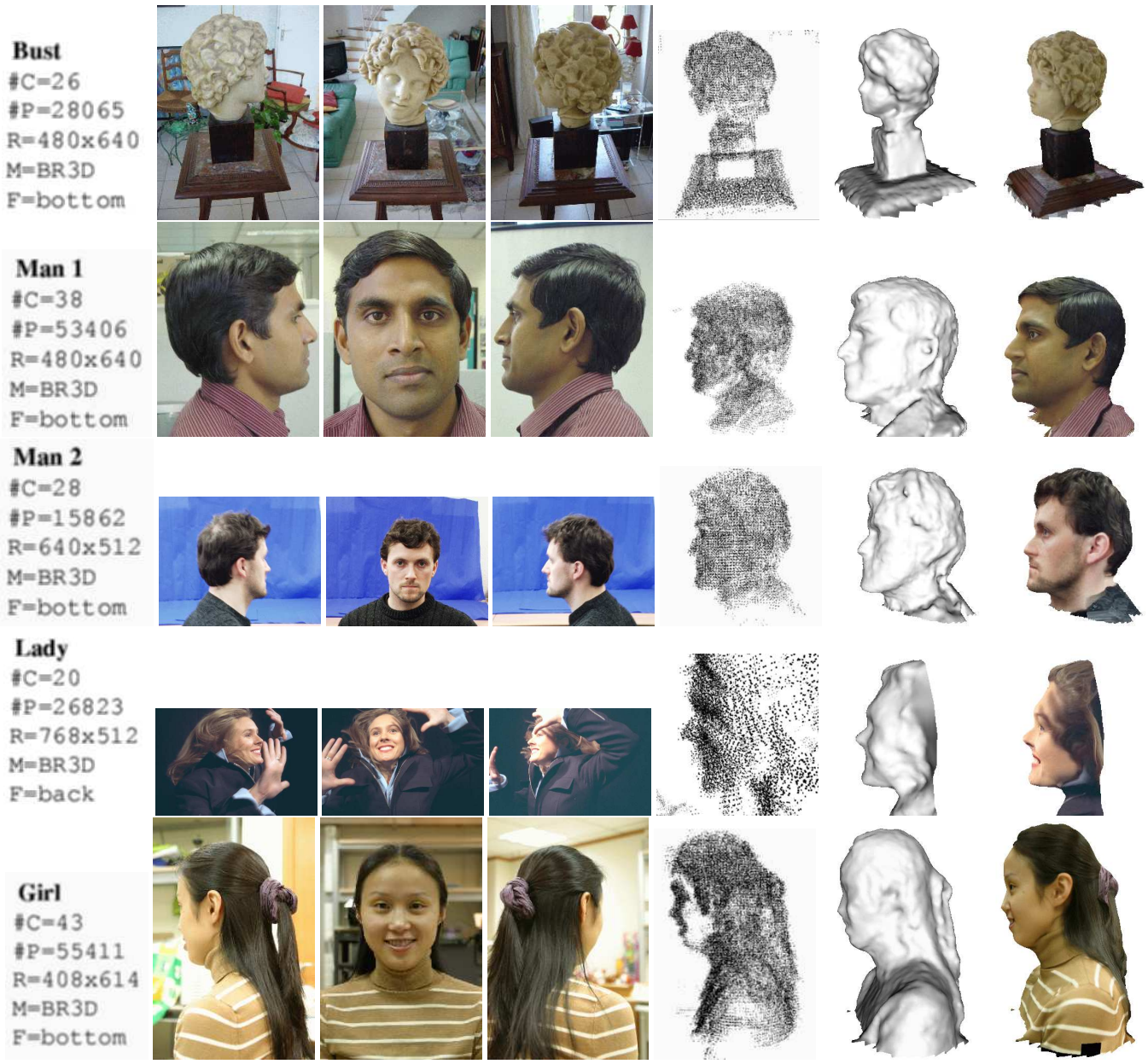

FIG. 5 – *Chaque ligne montre le r´esultat pour un exemple. De gauche `a droite: quelques d´etails exp´erimentaux, 3 images de la s´equence, points quasi-denses reconstruits, rendu de Gouraud pour la surface, rendu avec texture pour la surface. Pour les d´etails exp´erimentaux, $\#C$ est le nombre de cam´eras, $\#P$ est le nombre de points 3D, R les dimensions des images, et F la position du plan r´eduisant le domaine d'´evolution de la surface. Le temps de calcul pour l'´evolution de surface est d'environ 5 minutes sur un P4 `a 2.4 Ghz.*